TECHNICAL PAPER
CAA-TP-87-16

(21)

DTIC FILE COPY

AD-A196 126

DO BATTLES AND WARS HAVE A COMMON RELATIONSHIP BETWEEN CASUALTIES AND VICTORY?

NOVEMBER 1987

CONCEPTS ANALYSIS AGENCY
UNITED STATES ARMY

PREPARED BY
OFFICE, SPECIAL ASSISTANT FOR MODEL VALIDATION

US ARMY CONCEPTS ANALYSIS AGENCY
8120 WOODMONT AVENUE
BETHESDA, MARYLAND 20814-2797

CAA

DISCLAIMER

The findings of this report are not to be construed as an official Department of the Army position, policy, or decision unless so designated by other official documentation. Comments or suggestions should be addressed to:

Director
US Army Concepts Analysis Agency
ATTN: CSCA-MV
8120 Woodmont Avenue
Bethesda, MD 20814-2797

CAA-TP-87-16

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No 0704-0188 Exp. Date: Jun 30, 1986

1a. REPORT SECURITY CLASSIFICATION: UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S): CAA-TP-87-16	5. MONITORING ORGANIZATION REPORT NUMBER (S)
6a. NAME OF PERFORMING ORGANIZATION: US Army Concepts Analysis Agency — 6b. OFFICE SYMBOL (if applicable): CSCA-MV	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code): 8120 Woodmont Avenue, Bethesda, MD 20814-2797	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: CAA — 8b. OFFICE SYMBOL (if applicable): CSCA-MV	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
6c. ADDRESS (City, State, and ZIP Code): 8120 Woodmont Avenue, Bethesda, MD 20814-2797	10. SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO

11. TITLE (Include Security Classification)
Do Battles and Wars Have a Common Relationship Between Casualties and Victory?

12. PERSONAL AUTHOR(S)
Helmbold, Robert L.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Technical Paper	FROM Nov 87 TO Nov 87	1987 November	78

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This study examines whether a relation between casualties and victory in land combat battles, discovered in the course of earlier research, holds also for wars. It finds that this relation, or one similar to it, quite likely does hold for wars as well as battles. The paper also includes a brief exploration of the trend over historical time of the key quantities involved. It appears that the key variables involved have been quite stable from at least the early 1800s to the present day.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION: UNCLASSIFIED	
22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR — 83 Apr edition may be used until exhausted. All other editions are osolete

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

88 6 20 002

TECHNICAL PAPER
CAA-TP-87-16

DO BATTLES AND WARS HAVE A COMMON RELATIONSHIP BETWEEN CASUALTIES AND VICTORY?

November 1987

Accesion For	
NTIS CRA&I	☑
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

Prepared by

OFFICE, SPECIAL ASSISTANT FOR MODEL VALIDATION

US Army Concepts Analysis Agency
8120 Woodmont Avenue
Bethesda, Maryland 20814-2797

DEPARTMENT OF THE ARMY
US ARMY CONCEPTS ANALYSIS AGENCY
8120 WOODMONT AVENUE
BETHESDA, MARYLAND 20814-2797

REPLY TO
ATTENTION OF

CSCA-MVM

0 6 MAY 1988

MEMORANDUM FOR: Deputy Under Secretary of the Army (OR), Headquarters, Department of the Army, Washington, D.C. 20310

SUBJECT: Combat History Analysis Study Effort

1. The U.S. Army Concepts Analysis Agency (CAA) initiated the Combat History Analysis Study Effort (CHASE) in August 1984 to search for historically-based quantitative results for use in military operations research, concept formulation, wargaming, and studies and analyses.

2. Earlier CHASE work discovered a simple relation between casualties and victory in battle. This paper describes the results of an effort to determine whether this same relation also holds for wars as well as for battles. The results suggest that it, or some similar relation between casualties and victory, quite likely does hold for both battles and wars. However, two main difficulties prevented us from establishing that the relationship holds exactly for wars. First, not enough war data were available to establish a precisely definitive quantitative estimate of the relevant statistical parameters. Second, the available war data are not fully comparable with that on battles. For example, the war data give the entire national population at the start of the war, while the battle data give those military personnel actually present on the battlefield. Such lack of comparability tends to obscure any similarity of wars and battles regarding the relationship of casualties to victory.

3. Despite these limitations, the study has helped to clarify the range of validity for the relation between victory and casualties. Applications of these findings to wargames and studies are not necessarily simple or direct, but the results do (to give one example) suggest that gain or loss of ground by forces in contact may be more realistically represented by the relation of casualties to victory than by force ratio.

4. As is often the case with major original path-breaking research efforts, the illumination cast upon old issues brings several new ones to light. Accordingly, this Technical Paper is being provided to you now in the expectation that it will stimulate further research into the dynamics of combat and wars, as well as contribute information beneficial to those interested in the scientific and quantitative aspects of military wargaming and operations analysis. Questions or inquiries should be directed to the Office of Special Assistant for Model Validation, U.S. Army Concepts Analysis Agency (CSCA-MV), 8120 Woodmont Avenue, Bethesda, MD 20814-2797, (301) 295-1669.

E. B. Vandiver

E. B. VANDIVER III
Director

CAA	DO BATTLES AND WARS HAVE A COMMON RELATIONSHIP BETWEEN CASUALTIES AND VICTORY?	SUMMARY CAA-TP-87-16

THE REASON FOR PERFORMING THIS STUDY was to examine empirically the range of validity of a particular quantitative relation between the probability of victory in battles and the casualties on each side. This relation was discovered in the course of earlier research conducted at the US Army Concepts Analysis Agency (CAA). An empirical finding that the same relationship holds also for operations above the battle or tactical level would substantially strengthen the empirical basis for an inductive generalization that this relation is fundamental in determining victory in combat operations both at and above the tactical level. Since there are no quantitative data bases on combat operations at the campaign level, examining directly whether the relationship between casualties and victory holds for campaigns was not practicable. However, there are some quantitative data bases of wars that can be used for the purpose, although their data are not completely comparable to those for battles. Thus, using war data is a somewhat indirect approach to the study of whether the relation of victory to casualties found by earlier research to hold for battles holds also for campaigns and similar operations at the operational level. However, it was felt that, whatever its shortcomings, this indirect approach was the only currently feasible way to grapple empirically with the issue of whether or not this relationship between casualties and victory applies to combat operations above the tactical level.

THE PRINCIPAL FINDINGS are that this relation between casualties and victory, or some relation similar to it, quite likely does hold for wars as well as for land combat battles. It also appears that the key variables involved have been quite stable from the early 1800s to the present day.

THE MAIN ASSUMPTION is that the available war data are sufficiently error-free to allow at least a rough comparison to be made between them and the land combat battle data.

THE PRINCIPAL LIMITATIONS are two in number. First, not enough war data are available to establish a precisely definitive quantitative estimate of the relevant statistical parameters. Second, the available war data are not fully comparable with that on battles. For example, the war data give the entire national population at the start of the war, while the battle data give those military personnel actually present on the battlefield. Such lack of comparability tends to obscure any similarity of wars and battles regarding the relationship of casualties to victory.

THE SCOPE OF THE WORK is focused on examining whether the relation between casualties and victory in land combat battles, discovered in the course of earlier research, holds also for wars. The paper also includes a brief exploration of the trend over historical time of the key quantities involved, and identifies some issues that would make good topics for future research.

THE STUDY OBJECTIVE was to examine whether the relation between casualties and victory in battle, discovered in the course of earlier research, holds also for wars.

THE STUDY SPONSOR was the US Army Concepts Analysis Agency.

THE STUDY EFFORT was directed by Dr. Robert L. Helmbold, Office of Special Assistant for Model Validation, US Army Concepts Analysis Agency.

COMMENTS AND SUGGESTIONS may be sent to Director, US Army Concepts Analysis Agency, ATTN: CSCA-MV, 8120 Woodmont Avenue, Bethesda, MD 20814-2797.

Tear-out copies of this synopsis are at back cover.

CONTENTS

FIGURES

TABLES

PREFACE

This paper describes a portion of the work performed under the Combat History Analysis Study Effort (CHASE) project, which was begun in 1984 under the sponsorship of the US Army Concepts Analysis Agency (CAA). The overall objective of the CHASE project is to search for historically-based quantitative results for use in military operations research, concept formulation, wargaming, and studies and analyses. The CHASE project as a whole is oriented primarily to addressing the following essential elements of analysis (EEA):

1. Can the factors that have historically been most closely associated with victory in battle be identified?

2. What long-term trends can be detected in historical combat data?

3. Can the historical influence of air support on the outcome of battles be quantified?

4. What can be said about the factors influencing rates of advance in land combat?

5. What lessons were learned regarding the preparation of battle and engagement data bases for use in quantitative analyses?

The work described in this paper is limited to examining only selected aspects of EEAs 1 and 2. It was motivated by the thought that, should the relation between casualties and victory known to hold for battles hold also for wars, then it would strengthen the claim that this relation is a fundamental factor in determining victory in combat operations ranging in scale from battles through campaigns to wars. Since there are no adequate data bases on combat operations at the campaign level, examining directly whether the relationship between casualties and victory holds for campaigns is not practicable. Accordingly, it was felt that, whatever its shortcomings, the indirect approach adopted here was the only feasible way to grapple empirically with the issue of whether or not the relationship between casualties and victory applies to combat operations above the tactical level.

DO BATTLES AND WARS HAVE A COMMON RELATIONSHIP BETWEEN CASUALTIES AND VICTORY?

CHAPTER 1

EXECUTIVE SUMMARY

1-1. OBJECTIVE. This paper examines whether a quantitative relationship between casualties and the probability of victory, discovered in earlier research to hold empirically for land combat battles, holds also for wars.

1-2. BACKGROUND

a. Earlier research conducted at the US Army Concepts Analysis Agency (CAA) under the Combat History Analysis Study Effort (CHASE), as reported in Reference 1-1, discovered a quantitative empirical relationship between casualties and the probability of victory in battle. It also suggested the inductive generalization that this relationship is a fundamentally important, basic general property of combat operations. The relationship held for nearly all of the battles in the HERO data base (Ref 1-2), an extensive and detailed data base of 601 land combat battles that occurred between 1600 AD and 1973 AD. However, there were a few that did not fit the general pattern. Since most (though not all) of these anomalous cases were battles fought during World War II--specifically in the 1940-1949 decade--the phenomenon was called the World War II anomaly. Obviously, the presence of this anomaly tended to weaken claims that the relationship of casualties to victory was truly a fundamental general property of combat operations. Clearly, further investigations of the World War II anomaly were called for.

b. So far, two investigations bearing on the World War II anomaly have been conducted under the CHASE project. One of them is the subject of this Technical Paper and is dealt with at length below. The other was aimed at a careful review and reassessment of the anomalous battles to determine whether the data on them in the HERO data base of Reference 1-2 were affected by errors or inaccuracies, and in any case to obtain a completely independent assessment, both of the values reported in the HERO data base of Reference 1-2, and of the uncertainties surrounding those values. This work, described in Reference 1-3, was done by LFW Management Associates, Inc., under contract to CAA. Its authors concluded that "In virtually every case, the LFW Team's findings differ substantially from those determined by the authors of the HERO study. Not knowing the detailed processes employed in the HERO study, not having access to the final HERO study, and unaware of the reasons for the 61 battles being termed anomalous, the LFW Team can only reiterate that the figures the team has presented represent the closest possible approximation of the actual strengths in personnel, armament, casualties, and materiel losses." However, we have not yet had the opportunity to determine whether the differences in values reduces, exacerbates, or leaves essentially unchanged, the World War II anomaly. We hope in the near future to have an opportunity to undertake this examination.

c. The other investigation bearing on the World War II anomaly is the subject of this paper. It was motivated by the idea that, if the relationship between casualties and victory held for other data bases, this would support the view that the relationship is indeed general and probably fundamental. The CHASE progress report (Ref 1-1) did in fact confirm that the relationship held for another data base of land combat battles, as well as for some battles of both very ancient and very recent date, and those findings support the view that the relationship may be a very fundamental one that has held for land combat battles since antiquity. However, a finding that the relationship held for wars, and not just for battles, would be even more persuasive evidence that it is indeed a fundamentally important, basic general property of combat operations. Such a finding would also support the view that the World War II anomaly is primarily the result of errors in the data for battles of the 1940-1949 decade.

1-3. SCOPE. This paper examines whether the relationship between casualties and victory in land combat battles, discovered in earlier research, holds also for wars. It also examines briefly some historical trends involving the key factors and identifies some issues that would make good topics for future research.

1-4. LIMITATIONS

a. There are two principal limitations. One is that the war data provide too few data points for precise estimation of the parameters in the relationship between probability of victory and casualties. This limitation makes it difficult to determine with a high degree of assurance whether or not the war data follow the relationship between casualties and victory found to hold for land combat battles.

b. The other limitation is that the war data on casualties and strengths are not exactly compatible with the battle data on those quantities. To judge the impact of this, suppose for the sake of argument that wars do in fact follow the same relationship between casualties and victory as battles, when casualties in wars are taken into account in the same way as for battles. Then any lack of compatibility in the way casualties are taken into account will make the wars appear to follow a different relationship. We shall argue in Chapter 3 that the lack of compatibility greatly diminishes the prior expectation of finding the war data to follow exactly the same relationship as for battles. Accordingly, a finding that the wars do in fact follow roughly the same relationship would be rather remarkable.

1-5. TIMEFRAME. This paper uses battle data from the HERO data base of Reference 1-2, which includes 601 battles fought from 1600 AD to 1973 AD. It also uses war data collected by the University of Michigan's Correlates of War project as reported by Small and Singer (Ref 1-4), which includes data on 62 wars fought from 1823 AD to 1979 AD between national entities that satisfy certain criteria of statehood and thus qualify as members of what Small and Singer (Ref 1-4) call the "interstate system." The 62 wars are listed at Appendix C.

1-6. KEY ASSUMPTIONS. The key assumption is that the war data are sufficiently error-free to allow at least a rough comparison to be made between them and the battle data.

1-7. APPROACH

a. The approach adopted in this paper begins with a relation between victory in land combat battles and their casualties. This relation was obtained from the land combat battle data of the HERO data base (Ref 1-2) by using the logistic regression methods described in the CHASE progress report (Ref 1-1). This relation is described in detail in Chapter 2.

b. Then, using the University of Michigan Correlates of War project's (Ref 1-4) war data on casualties and prewar national population (and only that data), together with the relationship of victory to casualties and initial strengths for land battles, we attempt to determine or "predict" which sides in those wars were victorious. If this works well enough, then we have evidence that wars and battles follow the same relationship between casualties and victory.

1-8. FINDINGS. International wars (i.e., wars between what Small and Singer (Ref 1-4) includes as members of the "interstate system") are like land combat battles in at least the following respects.

a. The outcomes of wars having what Small and Singer (Ref 1-4) characterize as high confidence loss data are predicted quite well from the relation of victory to losses derived for battles. The outcomes of wars with what Small and Singer characterize as low confidence data are not well predicted by that relationship.

b. The association between predicted and actual winner for wars is much closer when the data confidence is high than when it is low.

c The fraction of wars won, lost, or drawn by the attacker is essentially the same as for battles.

d. The distribution of the defender's empirical advantage parameter for wars (ADV, as defined in paragraph 2-2) is close to that for battles.

e. The proportion of wars won by the attacker has not changed appreciably with time from the early 1800s to the present day. The same is true for battles.

f. The (defender's) ADV parameter for wars has not changed appreciably with time from the early 1800s to the present day. The same is true for battles.

g. What Small and Singer (Ref 1-4) characterize as the confidence level for data on wars has tended to decline with time from the early 1800s to the present day. Although it has not been definitely established that battle data follow a similar trend, this writer's informed judgment based on 30 years' experience working with quantitative battle data is that the confidence level for data on battles has also tended to decline with time over the same period.

h. The logistic regression functions and curves for the probability that the attacker wins versus ADV for the wars are qualitatively like those for land combat battles. A high degree of quantitative agreement is not anticipated for technical reasons discussed at greater length in Chapter 3. Nevertheless, for both wars and battles:

(1) Logistic regression intercepts are not significantly different from zero. Moreover, forcing a zero intercept value does not appreciably alter the estimated logistic regression slope.

(2) Small and Singer's (Ref 1-4) high and low data confidence levels tend to split their war data into two components exhibiting different logistic regression slopes. Steeper slopes are associated with high confidence data, and shallower ones with low confidence data. The same is true of battles.

1-9. PRINCIPAL OBSERVATIONS

a. The relationship of victories to casualties in wars is similar to that for land combat battles. Despite the lack of strict compatibility of the war and the battle data, and despite apparent differences between battles and wars, they share at least this relationship in common.

b. The key variables involved in this relationship appear to have been remarkably stable from at least the early 1800s to the present day. Since there is no empirical evidence that they will suddenly change in the forseeable future, it is rational to expect this relationship to persist.

c. Wars characterized by Small and Singer in Reference 1-4 as having high confidence casualty data follow the relationship between victory and casualties more faithfully than those with low confidence data. Accordingly, the apparent failure of some war data to follow the relationship exactly can reasonably be attributed to inaccurate or incomplete data, compounded by a lack of strict compatibility in the way casualties are treated in the war and in the battle data, and by the lack of a more extensive data base on wars.

d. In sum, the relationship of victory to casualties seems to be a fundamental one.

1-10. SUGGESTED TOPICS FOR FUTURE RESEARCH

a. Some research projects suggested by this work are mentioned below.

b. Bootstrap the logistic regressions of the war data (see References 1-5, 1-6, and 1-7). Compare the results to the logistic regression of battle data, or use some other "robust" method of logistic regression on the war data.

c. Obtain similar data on wars involving other political entities than those in Small and Singer's "system member" category. Determine whether or not they, too, are like land combat battles in their relation of victory to casualties.

d. See if enough data on wars can be obtained to place their ADV parameters on more nearly the same basis as ADV for battles. In particular, obtain enough good quality data on the losses per 1,000 armed forces personnel for wars to determine how closely they follow the same relationship of victory to casualties as do the ratios of battle death (BD) per unit population (i.e., BD/Pop ratios) given in Reference 1-4 and used in this paper.

e. Get the latest and best data available on BD/Pop ratios in wars involving "interstate system" members. Using it and the data in References 1-4 and 3-1, replicate (separately for each data set) the entire analysis. Investigate what differences in results are produced by the differences in data sets. Use the findings for historical criticism and for data base improvements.

f. Generate a data base of campaigns and see whether they, too, are like battles in their relation of victory to casualties.

g. Select from the data base of Reference 1-2 the longest and largest "battles"--some of which are the size of campaigns. Determine whether their relation of victory to casualties is the same as for the other battles in that data base, or whether it is "intermediate" between that for other battles and that for wars.

h. Perform for war data the same kinds of analyses as were done for land combat battles in References 4-1, 4-2, 4-3, 5-1, 5-2, 5-3, and 5-4. Explain the similarities and differences in the results for wars and battles.

i. Obtain data on historical naval and air battles and use it to replicate the entire analysis. Investigate what differences in results are produced by the different data sets.

CHAPTER 2

THE RELATION OF VICTORY TO CASUALTIES IN LAND COMBAT BATTLES

2-1. INTRODUCTION. This chapter describes the methods and the data used to determine the relation of victory to casualties in land combat battles and presents the results obtained.

2-2. THE RELATIONSHIP OF VICTORY TO CASUALTIES FOR LAND COMBAT BATTLES

a. The relationship is that the probability of an attacker victory is a logistic function of (the defender's) empirical advantage parameter, symbolized by ADV. We explain below what a logistic function is, and how ADV is determined from the battle data on strengths and losses.

b. A fairly general (multivariable multiple-response) logistic function is described in Appendix J of Reference 1-1. However, this paper uses only the simpler special case of a (univariate binary response) logistic function, defined by the equation

$$P(ADV) = EXP(a + b * ADV) / (1 + EXP(a + b * ADV)),$$

where P(ADV) is the probability that the attacker wins a battle in which the defender's advantage is ADV. The parameters a and b determine the exact form of the univariate logistic function on the right-hand side of the equation. They are called the logistic regression intercept and slope, respectively.

c. The (defender's) empirical advantage parameter is computed from the initial personnel strengths and losses in a battle as follows. Let x0 and x be the attacker's (and y0 and y be the defender's) initial and final personnel strengths. (These definitions tacitly assume that there is no replacement of personnel on either side. An adjustment for replacements is addressed in paragraph 2-3d, below.) Define A and D to be the attacker's and defender's fraction of personnel remaining at the end of the battle, i.e.,

$$A = x / x0,$$
$$D = y / y0,$$

and put

$$MU = SQR[(1 - A^2) / (1 - D^2)],$$

where SQR(z) is the square root of z. The (defender's) empirical advantage parameter is then defined to be

$$ADV = LOG(MU),$$

where LOG(z) is the natural logarithm of z. Reference 1-1 explains how the motivation for these definitions arises from a careful analysis of Lanchester's square law differential equations for attrition in combat. Reference 1-1 also shows that ADV is, both theoretically and empirically,

approximately equal to half the natural logarithm of the fractional exchange ratio, i.e.,

$$ADV \simeq (1/2) * LOG(FER),$$

with

$$FER = FX / FY,$$

where FX and FY are the attacker's and the defender's fractional losses, i.e.,

$$FX = 1 - A$$
$$FY = 1 - D.$$

2-3. FITTING THE LOGISTIC FUNCTION TO LAND COMBAT DATA

a. Introduction. The parameters a and b appearing in the logistic function for the probability of an attacker victory are selected to fit the battle data in Reference 1-2. The statistical method known as logistic regression, which is based on the theory of maximum likelihood estimation, is used to do the fitting. Logistic regression should not be confused with either logarithmic or linear regression, as it is quite different from both of these more widely-known regression methods. The logistic regression method is described in Appendix J of Reference 1-1 and in many statistical books and journal articles. This chapter concentrates on describing the data used and the results obtained, rather than on the statistical theory involved.

b. Choice of Battles to Use. Of the 601 battles in the data base of Reference 1-2, only the 427 non-World War II battles that occurred prior to 1940 and after 1949 were used in the logistic regression. The 1940-1949 decade was omitted because of the World War II anomaly and the resultant doubts about the quality of the data for that period of time, as mentioned in paragraph 1-2b.

c. Treatment of Drawn Battles. The data base of Reference 1-2 tabulates battle outcome as either a victory for the attacker, or a victory for the defender, or a draw. However, Table 1.6 of Reference 1-4 characterizes the outcomes of the war as either a victory for the attacker or for the defender. To make the battle data correspond to that for wars, in this paper, drawn battles are treated as a victory for the defender. Since only about 6 percent of the battles are scored as a draw, the values of a and b would be practically unaffected if draws were treated in some other way--such as ignoring them altogether, or allocating them randomly to victories by the attacker or by the defender.

d. Adjustment for Personnel Replacements. The equations presented in paragraph 2-2 above use the initial personnel strengths on each side. However, the data base of Reference 1-2 actually gives instead the total number of personnel "engaged" on each side. In most cases, this is the initial number. But in some cases, it is either an average daily strength or a total strength committed over the course of the battle. Accordingly, Reference 1-1 adjusted the values of x0, y0, x, and y to approximate an

estimated "initial strength equivalent." The same procedure, restated below, is also used in this paper. It is admittedly only a rough approximation to the effects of replacements over a lengthy battle. Fortunately, the logistic regression results are nearly the same whether the data are "adjusted" or not, partly because few of the battles in the data base satisfy the criteria for adjustment. In particular, only about 4 percent of the battles lasted at least 10 but less than 20 days, and only another 4 percent lasted longer than 20 days. The adjustment procedure used is as follows.

(1) If the battle duration is at least 10 but less than 20 days, the initial strengths are taken to be

$$x0 = (\text{Total engaged attacker personnel}) + Cx / 2$$

$$y0 = (\text{Total engaged defender personnel}) + Cy / 2$$

where (Total engaged ...) are the values actually listed in the data base of Reference 1-2, and Cx and Cy are the data base values for the attacker's and the defender's personnel losses.

(2) If the battle duration is at least 20 days, the initial strengths are taken to be

$$x0 = (\text{Total engaged attacker personnel}) + Cx$$

$$y0 = (\text{Total engaged defender personnel}) + Cy.$$

(3) In all cases the final strengths were taken to be

$$x = x0 - Cx$$

$$y = y0 - Cy,$$

This completes the process of adjusting the data to an "initial strength equivalent."

e. Values of the Logistic Regression Parameters. The values obtained from a logistic regression of winning sides versus the defender's empirical advantage parmeter (ADV), using only the non-WWII battle data, are:

$$\text{Logistic regression intercept} = a = - .02017$$

$$\text{Logistic regression slope} = b = -4.87764$$

Figure 2-1 shows a graph of the probability of an attacker victory determined using these logistic regression parameters.

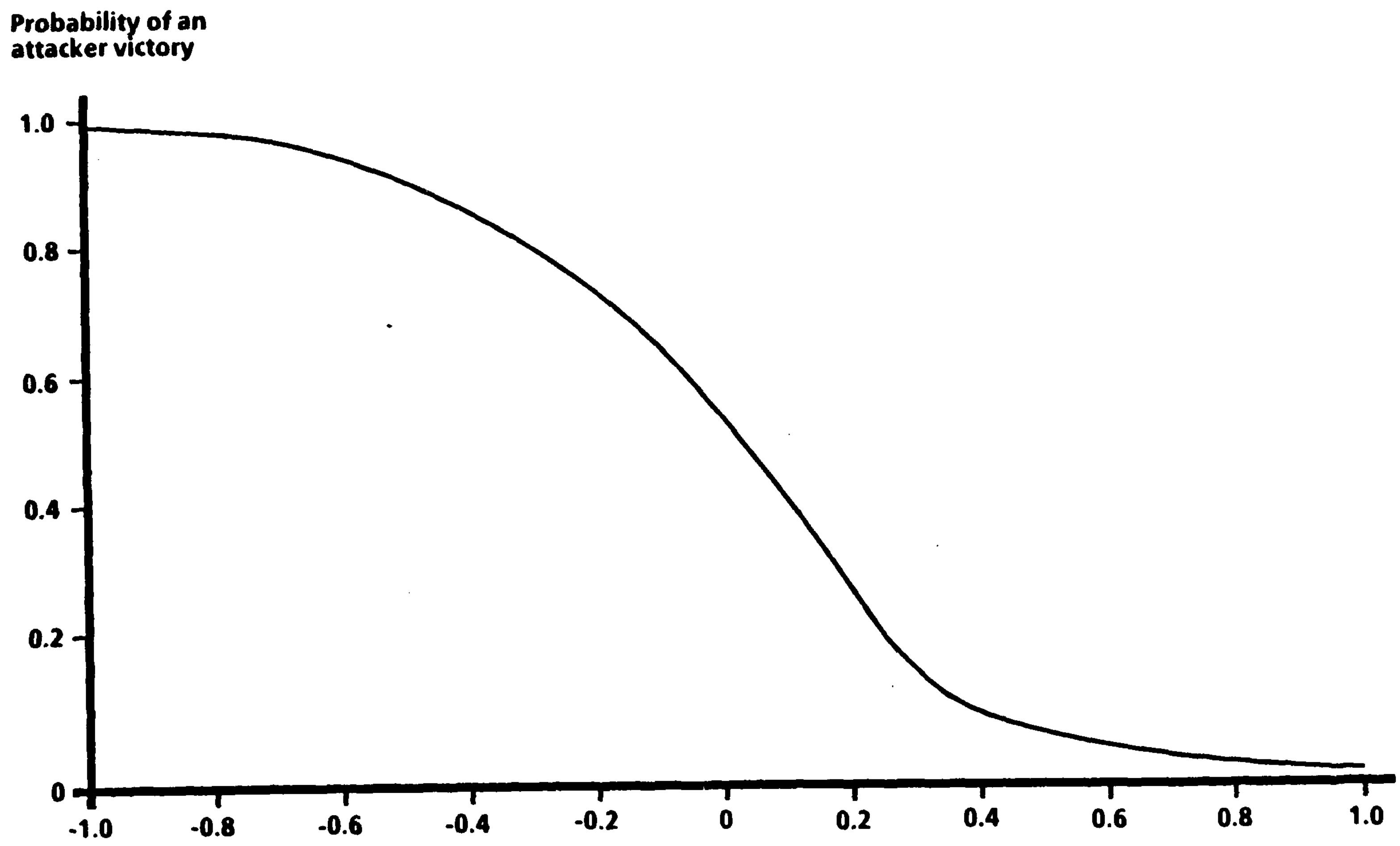

Figure 2-1. Probability of an Attacker Victory in Battle versus Defender's Advantage (logistic regression fit to non-WWII battle data)

CHAPTER 3

THE WAR DATA AND ITS COMPATIBILITY WITH THE BATTLE DATA

3-1. INTRODUCTION. This chapter begins by describing the war data used in this paper. It then relates the war data to the battle data and describes how the predictions of victory in war are made.

3-2. THE WAR DATA. The principal war data used in this paper are tabulated at Appendix C. They derive mainly from Table 11.6, "Initiation, Victory and Battle Death Ratios in Interstate Wars," of Reference 1-4. This table lists only those wars satisfying certain criteria, the most important of which are summarized below. In brief, the table includes only wars fought from 1823 to 1979 between "interstate system" members that incurred at least 1,000 battle deaths and which were not ties. Brief remarks on the compatibility of the war and the land combat battle data are included where appropriate. Our analyses of the war data also make use of the confidence levels reported on pages 73-74 of Reference 1-4. See Appendix C for a tabulation of the essential data derived from Reference 1-4 and used in this paper.

a. Table 11.6 of Reference 1-4 includes only 62 wars fought between 1823 and 1979. On the other hand, the 427 non-World War II battles used in this paper run from 1600 to 1973, with the exclusion of the decade 1940-1949.

b. Table 11.6 includes only wars between national entities that satisfy certain criteria of statehood, and hence qualify as members of the "interstate system." Table 2.1 of Reference 1-4 lists the members of the "interstate system" and shows how they changed with time. Thus, Table 11.6 of Reference 1-4 does not contain any civil wars. Nor does it contain any wars fought between (usually small, politically fragmented, and economically undeveloped or primitive) national entities that were not members of the "interstate system." Nor does it contain any (colonial or imperial) wars in which some "interstate system" member(s) fought against a nonmember national entity. Thus, the American Civil War is not listed in the table, nor is the British-Zulu war of 1879. On the other hand, the battle data include data on battles fought during some civil wars and a few colonial/imperialist wars, but not those fought entirely between irregular or unorganized military forces.

c. Table 11.6 does not include any wars in which fewer than 1,000 battle deaths were incurred altogether, including those on both sides. This was a more-or-less arbitrary decision on the part of the authors of Reference 1-4. They were interested in studying warlike phenomena and used 1,000 battle deaths as the threshold for distinguishing wars from minor incidents too small to qualify as "wars." On the other hand, although the battle data are not constrained in any formal way, they do focus on major or historically notable pitched battles involving larger forces and hence larger casualties. In particular, only about one-sixth of the battles had losses of 500 or less altogether, counting those on both sides.

d. Table 11.6 omits wars that were ties. Two wars that satisfied the other criteria (i.e., they were fought from 1823 to 1979 between interstate system members and incurred at least 1,000 battle deaths) are listed in Table 2.2 of Reference 1-4, but were excluded from Table 11.6 because--at least in the judgment of the authors of Reference 1-4--they were ties. These are the Korean War of 1950-1953 and the Israeli-Egyptian War of 1969-1970. On the other hand, the battle data lists battles that were drawn, and carefully distinguishes between attacker wins, defender wins, and drawn battles.

e. Three other interstate wars of the 67 listed in Table 2.2 of Reference 1-4 are also excluded from Table 11.6, presumably on the grounds that they were still in progress at the time Table 11.6 was compiled. Reference 1-4 lists them as the Vietnamese-Cambodian War of 1975-?, the Russo-Afghan War of 1979-?, and the Iran-Iraq War of 1980-?.

f. Pages 73-74 of Reference 1-4 give confidence levels on the battle death figures used in Table 11.6. Reference 1-4 eloquently expresses the major causes of uncertainty in those figures. "First, not all armed forces have been consistent in differentiating among dead, captured, missing, wounded, and deserting . . .. Second, there is the simple matter of accurate estimates, compounded by the fact that the size of a force may not be known with any accuracy even by its commanders. Third, there are the tactical reasons for exaggerating the enemy's losses and minimizing one's own. Finally, the archivists and historians who eventually sift through the reports and provide our basic sources of data may well suffer not only from a lack of statistical sophistication but even occasionally from personal and national biases of their own." Pages 73-74 of Reference 1-4 list the interstate wars under two levels of confidence, described as "high confidence," and as "somewhat lower." We will call them the high and the low confidence levels. The authors of Reference 1-4 note that, "Ironically enough, as the above lists indicate, the post-1945 period gave us more difficulty than the earlier period." On the other hand, Reference 1-2 provides no information on which battle data are more (or less) reliable than others.

3-3. COMPATIBILITY OF THE LOSS AND FORCE SIZE DATA ON WARS AND BATTLES

a. Table 11.6 of Reference 1-4 provides several data items for each of the wars listed in it. Our predictions of the victorious side in wars are based solely on the values in the column labeled "BD/Pop." The information in the column labled "Initiator Victor?" is used only for assessments of how accurate those predictions are. Paragraph 3-4 below describes how the predictions are made. Here we concentrate on the definition of information in the "BD/Pop" column, and on the extent to which it corresponds to the data in References 1-1 and 1-2 on land combat battles.

b. In Table 11.6 the initiator of a war is the side "whose battalions made the first attack in strength on their opponent's armies or territories." The initiator of a war corresponds to the attacker in a land combat battle, and the opponent in a war corresponds to the defender in a battle. We will call them the attacker (ATK) and defender (DEF), whether dealing with wars or battles.

c. The "BD/Pop" column in Table 11.6 of Reference 1-4 is the quotient obtained by dividing the opponent's battle deaths per unit prewar population by the corresponding item for the initiator. Thus, it corresponds roughly to the ratio FY/FX for battles, where FY is the defender's (and FX the attacker's) casualty fraction. Referring back to paragraph 2-2c above, we see that the BD/Pop ratio for a war roughly corresponds to the inverse of the fractional exchange ratio for a battle, i.e.,

$$\mathrm{BD/Pop} \leftrightarrow 1/\mathrm{FER} \leftrightarrow \mathrm{FY/FX}.$$

d. However, for the following reasons, this correspondence is far from exact.

(1) In the first place, the war data use "battle deaths," described in Reference 1-4 (page 70) as "combat-connected deaths of military personnel only." It is not clear exactly which "combat-connected" deaths are included in Reference 1-4's war data. Presumably they are not limited only to those killed in action, but include some ill-defined admixture of deaths due to illness, disease, nonbattle injuries, and died of wounds. On the other hand, the battle data use "battle casualties," described in Reference 1-2 as "The number of personnel killed, wounded, or missing (including prisoners) during the engagement. Does not (emphasis added) include personnel losses resulting from illness, disease, or nonbattle injuries." These battle losses are all directly attributable to combat action, but they include much more than just deaths. Thus, the figures on losses for wars and battles are at best only roughly compatible, even though both of them are for losses of military personnel only and do not include civilian losses.

(2) The battle data for the most part provide the initial forces engaged, although in a few cases the data base values were adjusted to approximate the initial figures as explained in paragraph 2-3d above. In contrast, the war data use the total prewar population, which of course includes women, children, the aged and infirm, and many others not fit for military service. Total prewar population does not relate to the military forces subject to experiencing "combat-connected deaths" nearly as closely as does the initial force engaged in a battle. Reference 1-4 (page 70) observes that "civilian deaths were quantitatively negligible in most international (as distinguished from civil) wars during our time span, except for the World Wars," so in nearly all cases, the total population was not significantly exposed to combat-connected losses. It would presumably have made the war data more compatible with the battle data if the combat-connected deaths had been related either to the active military forces or to the number of people subject to military service. For example, Table 4.2 of Reference 3-1 does give battle deaths per 1,000 armed forces for both sides in interstate wars, but the subsequent Reference 1-4 omits this information. Perhaps the authors decided that these data were not reliable enough to include in their later work. In any event, the size of the armed forces used in Reference 3-1 is only that at the start of the war, and this usually is considerably enlarged by both sides as the war progresses, especially for wars lasting at least as long as a few months. This growth in the size of the armed forces seriously clouds the validity of using just their size at the start of the war. In sum, the figures on initial force size for wars and battles are at best only roughly compatible.

3-4. PREDICTING VICTORY IN WARS

a. This paragraph describes how the war data were used with the battle relationship between victory and casualties to compare the relationship between the probability of victory and casualties found for battles with that for wars. For terminological convenience, we sometimes speak of the process as "predicting" the victorious side in a war, even though this is somewhat of a misnomer. Although the process has certain analogies with making predictions, our intent is simply to compare empirically the relationship between the probability of victory and casualties found for battles with that for wars.

b. The process begins by identifying the BD/Pop ratio given in Table 11.6 of Reference 1-4 with the inverse of the fractional exchange rate, i.e., we take

$$FER = 1 / (BD/Pop).$$

In view of the material in paragraphs 3-2 and 3-3, this relationship cannot be expected to be more than a rough approximation. That it may hold even roughly is due in no small part to the fact that the BD/Pop ratio is itself composed of ratios. Thus, if in war the number killed, wounded, and missing or taken prisoner is at least roughly proportional to the number of "combat-connected deaths," then the war losses become somewhat more nearly compatible with those used in computing the FER for battles. And if the number of personnel "engaged" in wars is at least roughly proportional to the total prewar population, then the number engaged for wars becomes somewhat more nearly compatible with the number used in computing FER for battles. In any event, the BD/Pop ratio (or, rather, its inverse) is more nearly compatible with the FER for battles than any other figure that is readily available. Any better estimates would require extensive, costly, and time-consuming original historical research.

c. The remaining steps are fairly straightforward. The attacker in a war is identified with its "initiator," in the terminology of Table 11.6 of Reference 1-4. The defender's empirical advantage value, ADV, for a war is estimated from its FER value by using the relation

$$ADV = (1 / 2) * LOG(FER),$$

which was obtained both theoretically and empirically in Reference 1-1. The probability that the attacker wins the war is estimated by substituting its ADV value into the logistic function fitted to the non-World War II battle data as described in paragraph 2-3. As stated in paragraph 2-3e, the logistic regression parameters obtained by fitting victory to ADV in battles are:

$$a = -.02017$$

$$b = -4.87764$$

Due to a lack of strict compatibility between the loss and force size data on wars and battles, we cannot expect this forecasting procedure to work perfectly. In fact, we should be astonished that it works at all.

CHAPTER 4

RESULTS

4-1. **INTRODUCTION.** This chapter describes the results of the predictions of victory in war, obtained as described in paragraph 3-4, and compares them to the actual victor. It also includes some additional comparisons of battles and wars, and provides a brief analysis of the trends of some of the key variables involved.

4-2. **TABULATION OF PREDICTION RESULTS.** Table 4-1 shows the number of wars won by the attacker and by the defender (or--in the terminology of Table 1-6 in Reference 1-4--by the "Initiator" and the "Opponent") for the variables listed below. The last column shows the observed fraction of wars actually won by the attacker. The key variables are:

a. The predicted probability P(ATKWIN) of an attacker victory, computed as described in paragraph 3-4. Four levels of P(ATKWIN) are used, viz, 0-0.25, 0.25-0.50, 0.50-0.75, 0.75-1.00.

b. The level of confidence in the loss data for the war, characterized as either high or low, as explained in paragraph 3-2g, Chapter 3.

Table 4-1. Number of Wars Won by Attacker and Defender

Confidence level	Predicted P(ATKWIN)	Actual winner			Observed fraction won by ATK
		ATK	DEF	Total	
High	0-0.25	1	5	6	0.200
	0.25-0.50	2	1	3	0.667
	0.50-0.75	4	1	5	0.800
	0.75-1.00	12	2	14	0.857
	(Subtotal)	(19)	(9)	(28)	(0.679)
Low	0-0.25	6	6	12	0.500
	0.25-0.50	1	0	1	1.000
	0.50-0.75	2	0	2	1.000
	0.75-1.00	14	5	19	0.737
	(Subtotal)	(23)	(11)	(34)	(0.676)
	Total	42	20	62	0.677

4-3. INITIAL OBSERVATIONS

a. As shown by the subtotal and total rows, the observed fraction of attacker victories is essentially the same for the high confidence data as for the low.

b. Also, the observed fraction of ATK wins seems to span a wider range for the high confidence data than for the low.

c. The predictions are "bold" in the sense that the totals for P(ATKWIN) values in the 0-0.25 and 0.75-1.00 range brackets strongly predominate. Thus, the predictions of which side is likely to win are seldom "wishy-washy."

d. Predictions of which side wins are more accurate for the high confidence data than for the low. The following discussion elaborates on this point.

(1) An extremely conservative approach would be to predict which side wins by flipping an unbiased coin. This method clearly could be expected to predict correctly the actual victor for about half of the wars.

(2) An improved but still conservative method for predicting the winner would be to observe that the attacker won most of the wars (about 67.7 percent of them, in fact), and so to predict which side wins by guessing in each case that the attacker would win. This method clearly would be expected to predict correctly the actual victor in about 67.7 percent of wars.

(3) Another method for predicting the winner would be to predict a defender win whenever P(ATKWIN) is less than 0.25 and an attacker win whenever P(ATKWIN) is more than 0.75. If the data confidence level is high, this predicts war outcomes correctly for 85 percent of the 20 wars in that category, as indicated in Table 4-2. But if the data confidence level is low, it results in a correct prediction rate of only 64.5 percent of the 31 wars in that category. Combining the high and low confidence levels results in a correct prediction rate of 72.5 percent of the 51 wars in the combined category.

(4) Similarly, the defender could be predicted to be the winner whenever P(ATKWIN) is less than 0.50, and the attacker could be predicted to be the winner otherwise. This results in a correct prediction rate of 78.6 percent of the 28 wars in the high confidence category, as indicated in Table 4-2. But if the data confidence level is low, it results in a correct prediction rate of only 64.7 percent of the 34 wars in the low confidence category. Combining the high and low confidence levels results in a correct prediction rate of 71.0 percent of the 62 wars in the data base.

Table 4-2. Proportion of Outcomes Predicted Correctly

Confidence level	Predicted P(ATKWIN) ranges	Number of outcomes			Proportion predicted correctly
		Predicted correctly	Predicted incorrectly	Total	
High	0-0.25 &0.75-1.00	17	3	20	0.850
	0-0-0.50 &0.50-1.00	22	6	28	0.786
Low	0-0.25 &0.75-1.00	20	11	31	0.645
	0-0.50 &0.50-1.00	22	12	34	0.647
High and low	0-0.25 &0.75-1.00	37	14	51	0.725
	0-0.50 &0.50-1.00	44	18	62	0.710

4-4. STATISTICAL ANALYSES OF THE PREDICTION RESULTS

a. The well-known chi-squared measure of association in contingency tables, a standard statistical technique explained in many textbooks and articles, is used to indicate the degree of association between the predicted values P(ATKWIN) and the observed frequency of an attacker victory in wars.

b. The chi-squared value for Table 4-1, taken as a whole, is 13.042 at 7 degrees of freedom. The probability of exceeding this value under the null hypothesis of no association between predicted and observed winner is 0.071, which is conventionally considered to be marginally significant statistically.

c. However, suppose Table 4-1 is divided into an upper half consisting of the high confidence data and a lower half consisting of the low confidence data.

(1) Then we find that the chi-squared value for the upper half, or high confidence data, is 9.595 at 3 degrees of freedom. The probability of exceeding this value under the null hypothesis of no association between predicted and observed winner is 0.022, which is conventionally considered to be significant statistically.

(2) We also find that the chi-squared value for the lower half, or low confidence data, is 3.459 at 3 degrees of freedom. The probability of exceeding this value under the null hypothesis of no association between predicted and observed winner is 0.326, which is conventionally considered to be definitely not significant statistically.

(3) Evidently the predicted winner is significantly associated with the actual winner for the high confidence levels, but not for the low one.

4-5. SOME OTHER COMPARISONS AND TRENDS

a. Table 4-3 shows the number of battles and wars the attacker won, lost, or drew. As can be seen, wars are very similar to battles in terms of the proportions won, lost, or drawn by the attacker.

Table 4-3. Number of Battles and Wars the Attacker Won, Lost, or Drew

Outcome	Battles		Wars	
	Number	Percent	Number	Percent[a]
ATKWIN	368	61.5	42	65.6
DRAW	34	5.7	2*	3.1
DEFWIN	196	32.8	20	31.3
Total	598	100.0	64*	100.0

[a]Two drawn wars are included with the 62 listed in Table 4-1; cf. paragraph 3-2d.

b. Figure 4-1 shows the cumulative empirical distribution of defender advantage for wars (estimated as described in paragraph 3-4c), together with a fitted normal distribution. As can be seen, the defender's advantage for wars is distributed approximately normally. As indicated in Figure 3-8 of Reference 1-1, ADV for battles also is distributed approximately normally. For the 62 wars in Table 11.6 of Reference 1-4, ADV has mean -0.384 and standard deviation 0.867. The corresponding values for battles are -0.300 and 0.741 (computed as half the values for LOG(FER) given in Table 3-6 of Reference 1-1, in accord with the formula in paragraph 3-4c above). Clearly, the distribution of ADV values for wars is close to that for battles.

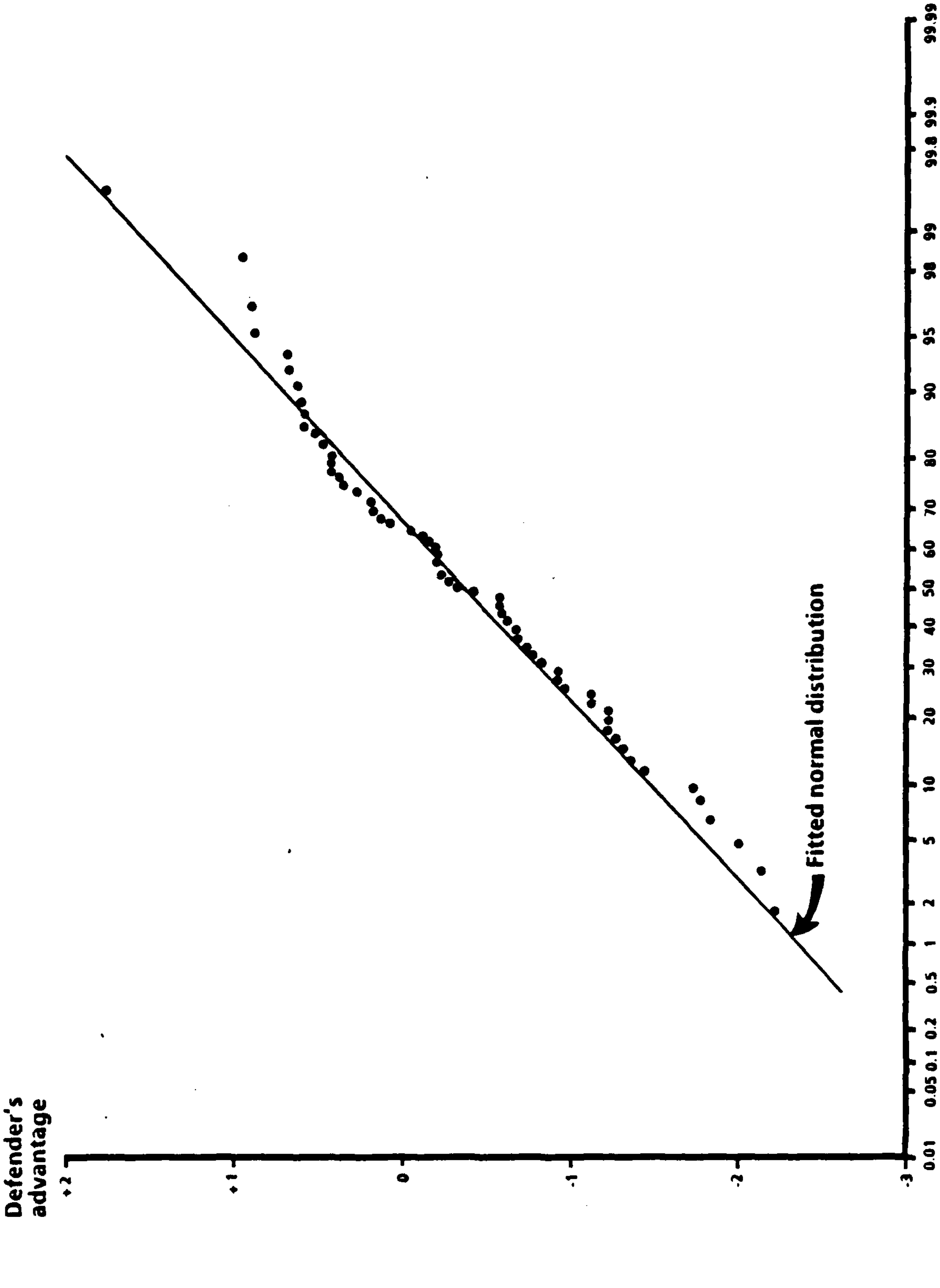

Figure 4-1. Distribution of Defender's Advantage for Wars

c. Table 4-4 shows the proportion of wars and battles won by the attacker over time. The time periods used here were selected to correspond to those used in Figure 3-2 and Table 3-3 of Reference 1-1.

Table 4-4. Proportion of Wars and Battles Won by Attacker Over Time

Years	Number of wars		Fraction won by attacker	Number of battles		Fraction won by attacker
	Won by attacker	Total		Won by attacker	Total	
1600-1699	ND[a]	ND	ND	36	48	0.75
1700-1799	ND	ND	ND	38	65	0.58
1800-1849	5	6	0.83	28	51	0.55
1850-1899	14	21	0.67	39	75	0.52
1900-1939	14	21	0.67	85	146	0.58
1940-1949	1	2	0.50	107	163	0.66
1950-1979	8	12	0.67	35	53	0.66
Total	42	62	0.68	368	601	0.61

[a]ND = no data.

(1) Treating the war data as a 5-by-2 contingency table of years-versus-winner yields a chi-squared value of 0.9841 at 4 degrees of freedom. The probability of exceeding this value under the null hypothesis of no association between winner and year is 0.922, which is conventionally considered to be definitely not significant statistically. A logistic regression of winner versus war date was performed for the data in Table 11.6 of Reference 1-4. For this logistic regression, 1900 was subtracted from all war dates to shift the "zero year" to 1900 AD. This yielded the following maximum likelihood estimates of the logistic regression parameters mentioned in paragraph 2-2b above (with standard error in parenthesis): a = 0.786 (0.28), b = -.00566 (0.0064). Clearly the logistic regression slope, b, is not statistically significantly different from zero, again indicating that the proportion of attacker victories for wars has been the same from the early 1800s to the present day.

(2) Treating the battle data as a 7-by-2 contingency table of years-versus-winner yields a chi-square value of 10.01 at 6 degrees of freedom. The probability of exceeding this value under the null hypothesis of no association between winner and year is 0.124, which is conventionally considered to be not significant statistically.

(3) So evidently, the proportion of attacker victories for wars as well as battles has been unchanged from the early 1800s to the present day.

d. A simple linear regression of (defender's) ADV versus date was performed for wars. The parameters are as indicated by the following equation:

$$ADV = a' + b' * (year - 1900) + e.$$

The maximum likelihood estimates of a' and b' (with standard errors shown in parentheses) are: $a' = -0.3840$ (0.11), $b' = 0.0000216$ (0.0026). The error e is distributed approximately normally with mean zero and standard deviation 0.867. Clearly the slope b' is not statistically significantly different from zero, so ADV does not depend on the date the war began. A similar result for battles is documented in References 4-1 through 4-3. So ADV is independent of date for wars as well as battles.

e. The war data's confidence level was logistically regressed against war date. The value 1900 was subtracted from all dates to shift the "zero year" to 1900 AD. The resultant maximum likelihood estimates of the regression parameters (with standard errors in parentheses) are $a = -0.1393$ (0.264), $b = -.0114$ (0.0062). Here a is not statistically significantly different from zero. But b is about 1.83 standard errors less than zero. The probability of observing a lesser value of b under the null hypothesis that b is zero is about 0.03, which is conventionally considered to be statistically significant. Thus there appears to have been a tendency for the probability of finding high confidence data on a war to decline as its date increases--at least for the period that begins in the early 1800s and ends with the present time. Using the logistic regression parameters, we estimate that the probability of finding high confidence data for a war declined from about 0.68 in 1820, to 0.58 in 1860, to 0.47 in 1900, to 0.36 in 1940, and to 0.26 in 1980. There are no comparable quantitative studies of battle data reliability versus battle date. However, this writer has for some time now held the opinion--based on long, detailed and extensive practical experience with several data bases on land combat battles--that battle data reliability does not necessarily increase with battle date. In fact, the most accurate data on land combat battle may be from the Napoleonic and American Civil Wars, since they have been so carefully studied for so long by so many highly qualified military historians.

4-6. LOGISTIC REGRESSIONS OF WINNER VERSUS ADVANTAGE

a. This paragraph reports the results of some logistic regressions of the winner versus ADV. In all of these logistic regression computations, draws are counted as defender wins. Logistic regression techniques are more sensitive, but less robust, than the methods used earlier in this chapter. Thus, the logistic regression results are more likely to be influenced adversely by a few gross errors in the data, or by any lack of strict

compatibility of the war with the battle data. Hence, we do not anticipate the logistic regression parameters fitted to the war data to be in very good quantitative agreement with those fitted to the battle data. War data that are accurate and more nearly compatible with battle data might very well produce logistic regression parameters that agree quite well with those for battles.

b. Table 4-5 presents the logistic regression parameters fitted to various data sets by the method of maximum likelihood, together with the approximate standard errors in those estimates. The WWII data set consists of all of the battles in Reference 1-2 that started between 1 January 1940 and 31 December 1949. The non-WWII data set consists of all other battles in Reference 1-2. (Battles in Reference 1-2 that have insufficient data to compute their ADV parameter are omitted from both of these data sets.) The high confidence war data set consists of those wars listed in Table 11.6 of Reference 1-4 whose data is characterized as high confidence. The low confidence war data set consists of the other wars from Table 11.6 of Reference 1-4. The following analysis largely ignores the logistic regression intercept values (a) and focuses instead on the slopes (b). The reasons for this are as stated immediately below.

Table 4-5. Logistic Regression Parameters Fitted to Various Data Sets

Data set	Number of data items	Maximum likelihood estimates		Standard errors	
		a	b	SE(a)	SE(b)
Non-WWII battles	427	-0.0202	-4.8776	0.1374	0.5043
High confidence wars	28	0.3121	-2.0733	0.4800	0.9176
WW II battles	158	0.4646	-0.6543	0.1993	0.2612
Low confidence wars	34	0.5890	-0.5371	0.3862	0.4304

(1) Different intercept values simply slide the logistic curve left and right, but do not change its shape.

(2) Except for the WWII battle data set, the logistic regression intercept is not significantly different from zero (differs from zero by less than two standard errors). We elect to treat this as a fluke or nonsignificant characteristic of the WWII data set.

(3) Except for the WWII battle data set, forcing a zero value of the intercept does not appreciably change the estimated slope. For the WWII data set, forcing a zero intercept yields an estimated slope of about -0.9662. We elect to treat this as a fluke or nonsignificant characteristic of the WWII data set.

c. Figure 4-2 shows, in graphical form, essentially the same information presented in Table 4-5. Observe that:

(1) The low confidence war data have logistic regression slopes similar to those for the WWII battle data.

(2) The high confidence war data have logistic regression slopes much steeper than those for either the WWII battles or the low confidence wars. Figure 4-3 shows that the logistic regression function fitted to the high confidence war data is qualitatively more like that for the non-WWII battles than for either the WWII battles or the low confidence war data.

(3) The high and the low confidence levels tend to split the war data into two components whose logistic regression slopes are noticeably different. This is qualitatively analogous to the way the battle data fission into WWII and non-WWII components. Reference 1-1 presented good reasons for believing that the WWII battle data are less reliable than the non-WWII battle data, which also supports the analogy.

d. These results indicate that the war data are qualitatively similar to the land combat battle data with respect to the relation of victory to casualties.

4-6. CONCLUSION. This chapter has shown that the war data are similar to the battle data in several important respects and has presented some findings on trends. The principal results are summarized in paragraph 5-2, Chapter 5.

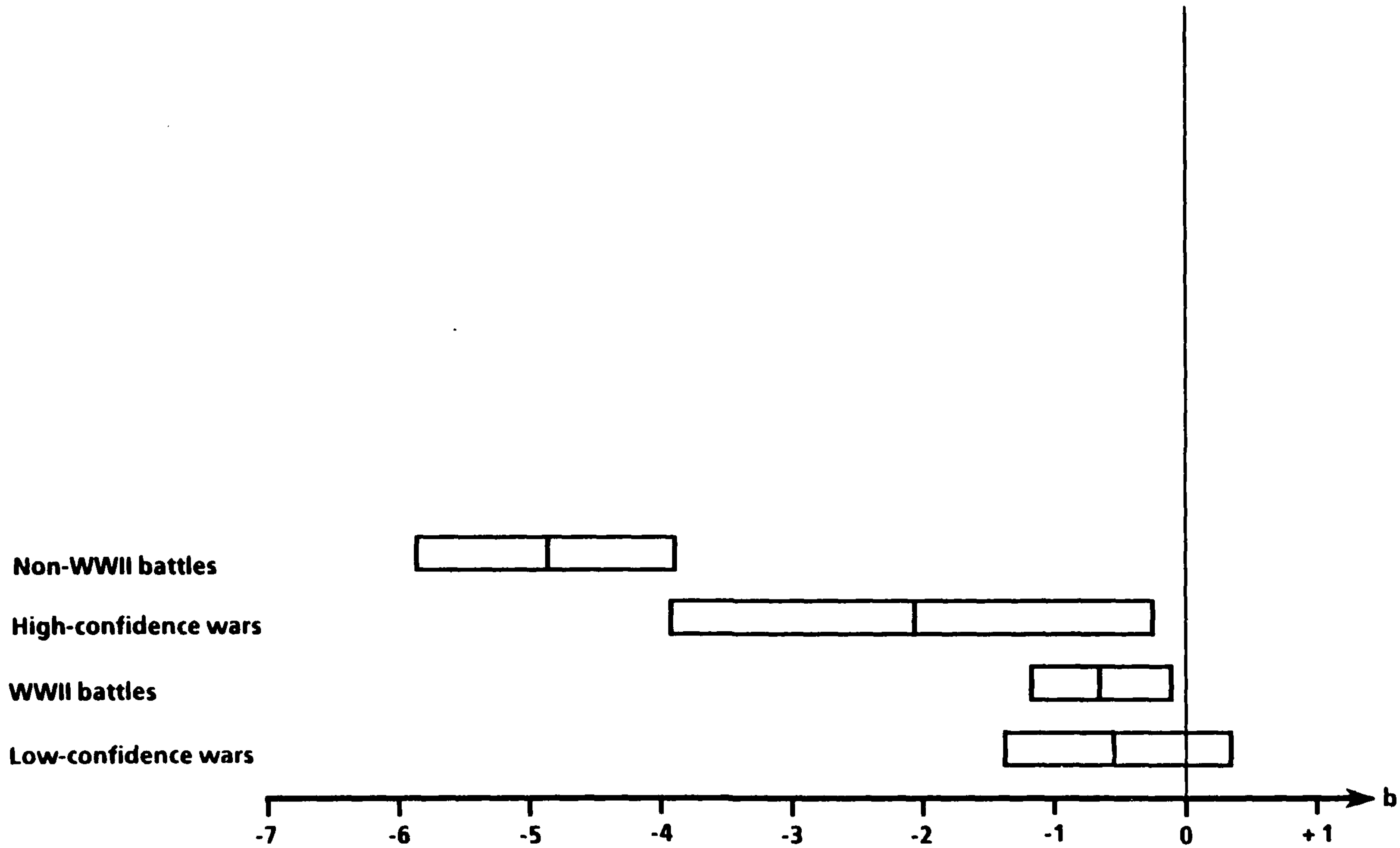

Figure 4-2. Estimated Values and (+-2 Standard Error) Tolerance Bands for Logistic Regression Slopes, b

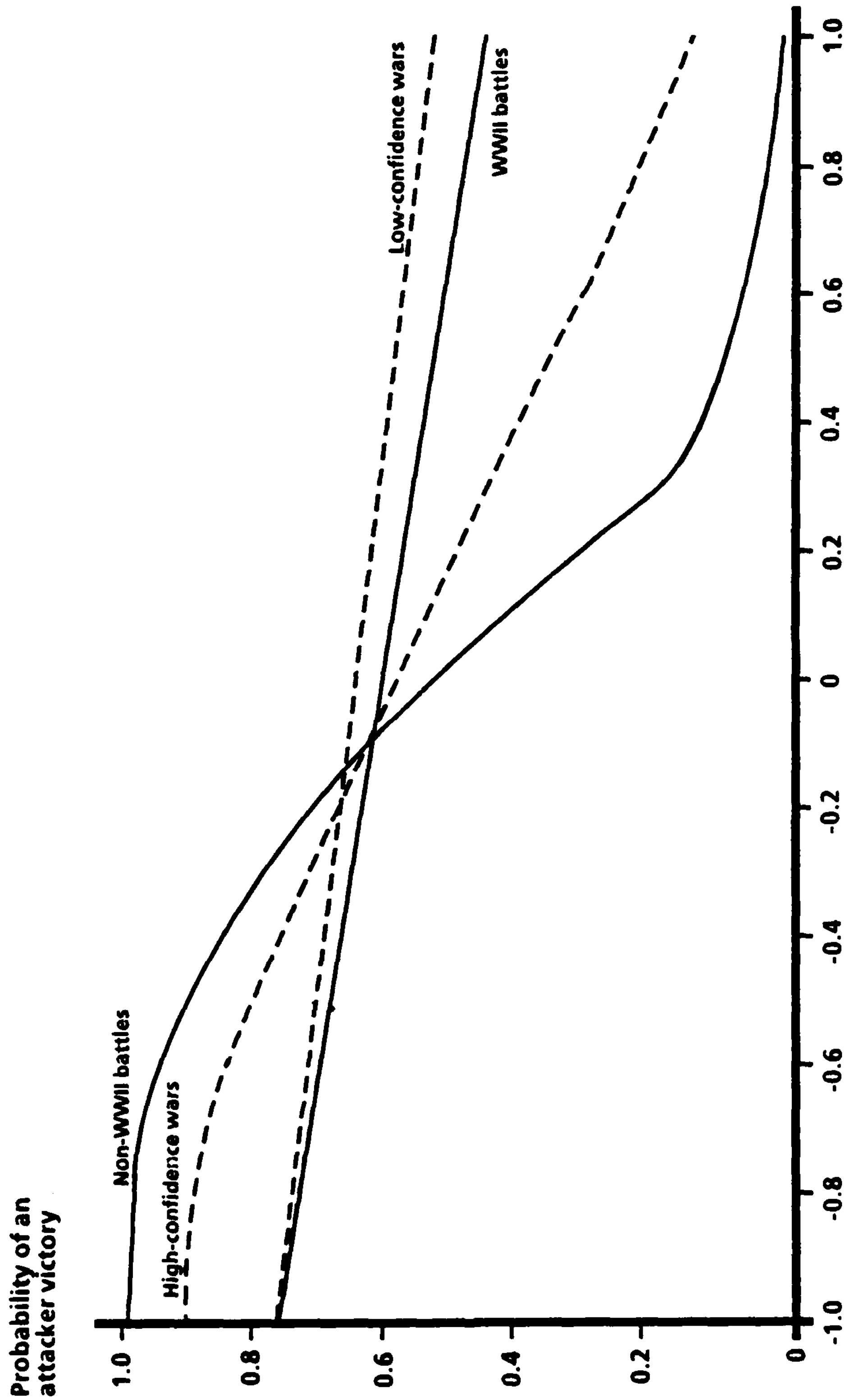

Figure 4-3. Fitted P(ATKWIN) Curves for Various Data Sets

CHAPTER 5

SUMMARY OF FINDINGS, AND OBSERVATIONS

5-1. INTRODUCTION. This chapter first summarizes the findings, then presents the conclusions and observations. It ends with some suggested topics for future research.

5-2. SUMMARY OF FINDINGS. Wars between members of the "interstate system" are like land combat battles in at least the following respects.

a. The outcomes of wars having what Small and Singer (Ref 1-4) characterize as high confidence loss data are predicted quite well by the relation of victory to losses derived for battles. The outcomes of wars with what Small and Singer characterize as low confidence data are not well predicted by that relationship.

b. The association between predicted and actual winner for wars is much closer when the data confidence is high than when it is low.

c. The fraction of wars won, lost, or drawn by the attacker is essentially the same as for battles.

d. The distribution of the defender's empirical advantage parameter for wars (ADV, as defined in paragraph 2-2) is close to that for battles.

e. The proportion of wars won by the attacker has not changed appreciably with time from the early 1800s to the present day. The same is true for battles.

f. The (defender's) ADV parameter for wars has not changed appreciably with time from the early 1800s to the present day. The same is true for battles.

g. What Small and Singer (Ref 1-4) characterize as the confidence level for data on wars has tended to decline with time from the early 1800s to the present day. Although it has not been definitely established that battle data follow a similar trend, this writer's informed judgment is that the confidence level for data on battles has also tended to decline with time over the same period.

h. The logistic regression functions and curves for the probability that the attacker wins versus ADV for the wars are qualitatively like those for land combat battles. A high degree of quantitative agreement is not anticipated for the technical reasons discussed in Chapter 3. Nevertheless, for both wars and battles:

(1) Logistic regression intercepts are not significantly different from zero. Moreover, forcing a zero intercept value does not appreciably alter the estimated logistic regression slope.

(2) Small and Singer's high and low confidence characteristics tend to split the war data into two components exhibiting different logistic regression slopes. Steeper slopes are associated with high confidence data, and shallower ones with low confidence data. The same is true of battles.

5-3. PRINCIPAL OBSERVATIONS

a. The relationship of victories to casualties in wars is similar to that for land combat battles. Despite the lack of strict compatibility of the war and the battle data, and despite apparent differences between battles and wars, they share at least this relationship in common.

b. The key variables involved in this relationship appear to have been remarkably stable from at least the early 1800s to the present day. Since there is no empirical evidence that they will suddenly change in the foreseeable future, it is rational to expect this relationship to persist.

c. Wars characterized by Small and Singer in Reference 1-4 as having high confidence casualty data follow the relationship between victory and casualties more faithfully than those with low confidence data. Accordingly, the apparent failure of some war data to follow the relationship exactly can reasonably be attributed to inaccurate or incomplete data, compounded by a lack of strict compatibility in the way casualties are treated in the war and in the battle data, and by the lack of a more extensive data base on wars.

d. In sum, the relationship of victory to casualties seems to be a fundamental one.

5-4. OTHER OBSERVATIONS

a. Since the relationship between victory and ADV for wars is the same as that for battles, it is reasonable to conjecture that the same relationship holds for campaigns.

b. Since errors in the data seriously affect the logistic regression results, it is reasonable to conjecture that at least a part of the quantitative difference between the logistic regression parameters computed for high confidence wars and battles is due to errors that affect even the "high confidence" war data. The findings tend to support the hypothesis that the difference in logistic regressions for WWII and non-WWII battles is also due in large part to errors in the data on battles in the WWII data set. In other words, these findings support the view that the World War II anomaly is primarily the result of errors in the data for battles of the 1940-1949 decade.

c. Incompatibility of the war and the battle data may account for the remaining quantitative differences between the logistic regression parameters computed for high confidence wars and battles.

5-5. SUGGESTED TOPICS FOR FUTURE RESEARCH. Some research projects suggested by this work are mentioned below.

a. Bootstrap the logistic regressions of the war data. Compare the results to the logistic regressions of battle data, or use some other "robust" method of logistic regression on the war data.

b. Obtain similar data on wars involving other than "system member" participants. Determine whether or not they, too, are like land combat battles in their relation of victory to casualties.

c. See if enough data on wars can be obtained to place their ADV parameters on more nearly the same basis as ADV for battles. In particular, obtain enough good quality data on the losses per 1,000 armed forces personnel for wars to determine how closely it follows the same relationship of victory to casualties as for the BD/Pop ratios given in Reference 1-4 and used in this paper.

d. Get the latest and best data available on BD/Pop ratios in wars involving "interstate system" members. Using it and the data in References 1-4 and 3-1, replicate (separately for each data set) the entire analysis. Investigate what differences in results are produced by the differences in data sets. Use the findings for historical criticism and for data base improvements.

e. Generate a data base of campaigns and see whether they, too, are like battles in their relation of victory to casualties.

f. Select from the data base of Reference 1-2 the longest and largest "battles"--some of which are the size of campaigns. Determine whether their relation of victory to casualties is the same as for the other battles in that data base, or whether it is "intermediate" between that for other battles and that for wars.

g. Perform for war data the same kinds of analyses as were done for land combat battles in References 4-1, 4-2, 4-3, 5-1, 5-2, 5-3, and 5-4. Explain the similarities and differences in the results for wars and battles.

h. Obtain data on historical naval and air battles and use it to replicate the entire analysis. Investigate what differences in results are produced by the different data sets.

APPENDIX A

STUDY CONTRIBUTORS

1. STUDY TEAM

Study Director

Dr. Robert L. Helmbold, Office, Special Assistant for Model Validation

2. PRODUCT REVIEW BOARD

Mr. Carl A. Steinhagen, Chairman

Dr. Ralph E. Johnson

MAJ Eric A. Vonderscheer

Mr. Giles Mills III

APPENDIX B

REFERENCES

1-1. Combat History Analysis Study Effort (CHASE): Progress Report for the Period 1984-June 1985, CAA-TP-86-2, August 1986. AD F860 122

1-2. Analysis of Factors That Have Influenced the Outcomes of Battles and Wars: A Data Base of Battles and Engagements, Vols I-VI, CAA-SR-84-6, September 1984. (Available under AD numbers AD-B086 797L, AD-B087 718L, AD-B087 719L, AD-B087 720L, AD-B087 720L, AD-B087 721L, AD-B087 722L).

1-3. Independent Review/Reassessment of Anomalous Data, LFW Management Associates, Inc., prepared for US Army Concepts Analysis Agency under Contract No. MDA903-86-C-0396.

1-4. Resort to Arms: International and Civil Wars, 1816-1980, Melvin Small and J. David Singer, SAGE Publications, Inc., Beverly Hills, CA, 1982.

1-5. The Jacknife, the Bootstrap and Other Resampling Plans, Bradley Efron, Society for Industrial and Applied Mathematics, 1982.

1-6. A Leisurely Look at the Bootstrap, The Jacknife and Cross-Validation, Bradley Efron and Gail Gong, The American Statistician 37, 36-48, 1983.

1-7. Computer Intensive Methods in Statistics, Persi Diaconis and Bradley Efron, Scientific American, May 1983, 116-130.

3-1. The Wages of War, 1816-1965: A Statistical Handbook, J. David Singer and Melvin Small, John Wiley & Sons, Inc., NY, 1972.

4-1. Historical Data and Lanchester's Theory of Combat, Robert L. Helmbold, Combat Operations Research Group Staff Paper, CORG SP-128, HQ, US Continental Army Command, 1961. (AD 480 975L)

4-2. Historical Data and Lanchester's Theory of Combat: Part II, Robert L. Helmbold, Combat Operations Research Group Staff Paper, CORG SP-190, HQ, US Army Combat Developments Command, August 1964. (AD-480 109L)

4-3. Integrated Warfare--Representing the Decision Processes Within Computer Simulations of Combat, Technical Paper CAA-TP-85-1, US Army Concepts Analysis Agency, Bethesda, MD, January 1985.

5-1. Probability of Victory in Land Combat as Related to Force Ratio, Robert L. Helmbold, RAND Paper P-4199, The RAND Corporation, Santa Monica, CA, October 1969.

5-2. Decision in Battle: Breakpoint Hypotheses and Engagement Termination Data, Robert L. Helmbold, RAND Report R-772-PR, The RAND Corporation, June 1971.

5-3. Some Observations on the Choice of Exponent in Lanchester's Law, Robert L. Helmbold, presented to the Ad Hoc Working Group on the Theory of Combat, co-sponsored by the Defense Nuclear Agency (DNA) and the Naval Postgraduate School (NPGS), held at NPGS, Monterey, CA, 1979.

5-4. The Influence of the Numerical Strength of Engaged Sides on Their Losses, M. Osipov, a five-part series of articles (in Russian) published in Voenniny Sbornik, issues 6 (June) through 10 (October), 1915.

APPENDIX C

DATA TABLE

C-1. Table C-1 presents the data on wars between members of the "interstate system" that were used in the remainder of this paper. It is based mainly on the material provided in Table 11.6 of Reference 1-4. The following paragraphs explain in each case where additional information provided in Reference 1-4 is used.

C-2. Column one is a sequence or line number assigned by CAA.

C-3. Column two gives the (defender's) advantage parameter, ADV, estimated from the BD/Pop values given in Table 11.6 of Reference 1-4 as described in paragraph 3-4.

C-4. Column three gives P(ATKWIN), the probability that the attacker wins the war, computed by substituting the war's ADV parameter into the logistic function fitted to the non-WWII land battle data. That logistic function is described in paragraph 2-3e and in Figure 2-1. A similar result would be obtained by reading the P(ATKWIN) value for the war's ADV parameter from Figure 2-1.

C-5. Column four gives the war's actual victor, using the notation I for a win by the initiator (i.e., the attacker) and 0 for a win by his opponent (i.e., the defender).

C-6. Column five gives the loglikelihood of the war's observed outcome, relative to the P(ATKWIN) values computed from the logistic function fitted to the non-WWII land battle data as described in paragraph 2-3e and in Figure 2-1. The loglikelihood of a war outcome is, by the standard statistical definition, given by

$$\text{LOGLIKELIHOOD} = \text{LOG}(\text{P(ATKWIN)}), \text{ if outcome} = \text{I},$$
$$\text{LOGLIKELIHOOD} = \text{LOG}(\text{P(DEFWIN)}), \text{ if outcome} = 0,$$

where, of course,

$$\text{P(DEFWIN)} = 1 - \text{P(ATKWIN)}.$$

C-7. Column six gives the sequence or index number that is used in Reference 1-4 for the war.

C-8. Column seven gives the name given in Table 4-2 of Reference 1-4 for the war.

C-9. Column eight gives the year the war started according to Table 4-2 of Reference 1-4.

C-10. Column nine, the last column, gives the level of confidence (high or low) on the loss data for the war. These confidence levels are provided on pages 73-74 of Reference 1-4, as described in paragraph 3-2g.

Table C-1. Data from Reference 1-4
(page 1 of 3 pages)

CAA index no	ADV	P(ATKWIN)	Actual winner	LOGLIHOD	Ref 1-4's seq no	War name	Start year	Data confidence
1	0.688	.9657	I	-.03495	1	Franco-Spanish	1823	H
2	0.666	.9619	I	-.03882	4	Russo-Turkish	1828	L
3	0.199	.7216	I	-.3263	7	Mexican-American	1846	H
4	-0.602	.04943	O	-.0507	10	Austro-Sardinian	1848	H
5	1.144	.9962	I	-.003837	13	First Scheleswig-Holstein	1848	H
6	2.161	1.0000	I	-.00002692	16	Roman Republic	1849	H
7	1.236	9975	I	-.002458	19	La Plata	1851	L
8	-0.075	.4042	I	-.9058	22	Crimean	1853	H
9	2.215	1.0000	I	-.0000207	25	Anglo-Persian	1856	L
10	-0.164	3055	O	-.3645	28	Italian Unification	1859	H
11	0.580	9432	I	-.05852	31	Spanish-Moroccan	1959	L
12	1.362	9987	I	-.001329	34	Italo-Roman	1860	H
13	0.196	7183	I	-.3309	37	Italo-Sicilian	1860	H
14	0.930	.9892	O	-4.5293	40	Franco-Mexican	1862	L
15	0.902	.9876	I	-.01247	43	Ecuador-Columbian	1863	H
16	1.832	.9999	I	-.0001341	46	Second Schleswig-Holstein	1864	H
17	-1.753	.0001893	O	-.0001893	49	Lopez	1864	L
18	1.105	.9954	O	-5.3727	52	Spanish-Chilean	1865	H
19	0.203	7249	I	-.3218	55	Seven Weeks	1866	H
20	-0.586	05333	O	-.0548	58	Franco-Prussian	1870	H
21	0.767	9764	I	-.02388	61	Russo-Turkish	1877	L
22	0.316	.8204	I	-.1979	64	Pacific	1879	L
23	-0.471	08976	I	-2.4106	67	Sino-French	1884	L

Table C-1. Data from Reference 1-4
(page 2 of 3 pages)

CAA index no	ADV	P(ATKWIN)	Actual winner	LOGLIHOD	Ref 1-4's seq no	War name	Start year	Data confidence
24	-0.388	.1285	O	-.1376	70	Central American	1885	H
25	-0.886	.01285	I	-4.3546	73	Sino-Japanese	1894	L
26	-0.693	.03226	O	-.03279	76	Greco-Turkish	1897	H
27	0.689	.9659	I	-.03474	79	Spanish-American	1898	H
28	0.165	.6863	I	-.3764	82	Boxer Rebellion	1900	L
29	-0.886	.01285	I	-4.3546	85	Russo-Japanese	1904	L
30	0.235	.7551	I	-.2809	88	Central American	1906	H
31	-0.112	3625	I	-1.0146	91	Central American	1907	H
32	1 431	9991	I	- 000947	94	Spanish-Moroccan	1909	L
33	0.591	9459	I	-.05559	97	Italo-Turkish	1911	L
34	-0.674	.03539	I	-3.3415	100	First Balkan	1912	L
35	-0.586	.05333	O	-.0548	103	Second Balkan	1913	L
36	-0.367	.1406	O	-.1515	106	World War I	1914	L
37	0.582	.9436	O	-2.8746	109	Russo-Polish	1919	L
38	0.621	.9529	I	-.04825	112	Hungarian-Allies	1919	L
39	-0.636	.0421	O	-.04301	115	Greco-Turkish	1919	L
40	0.723	.9709	I	-.0295	118	Sino-Soviet	1929	L
41	-0.200	2696	I	-1.311	121	Manchurian	1931	L
42	-0.272	2061	I	-1.5794	124	Chaco	1932	H
43	1.321	9984	I	- 001625	127	Italo-Ethiopian	1935	H
44	-0.388	1285	I	-2.0516	130	Sino-Japanese	1937	L
45	0.048	.5529	O	- 8049	133	Changkufeng	1938	H
46	-0.886	.01285	O	- 01293	136	Nomonhan	1939	H
47	-0.378	.1345	O	-.1445	139	World War II	1939	L

Table C-1. Data from Reference 1-4
(page 3 of 3 pages)

CAA index no	ADV	P(ATKWIN)	Actual winner	LOGLIHOD	Ref 1-4's seq no	War name	Start year	Data confidence
48	1.777	.9998	I	-.0001754	142	Russo-Finnish	1939	L
49	-0.511	.07504	I	-2.5898	145	Franco-Thai	1940	L
50	1.274	.9980	O	-6.1938	148	Palestine	1948	L
51	0.954	.9904	I	-.009676	154	Russo-Hungarian	1956	L
52	1.996	.9999	I	-.0000603	157	Sinai	1956	H
53	0.199	.7216	I	-.3263	160	Sino-Indian	1962	L
54	1.213	.9973	O	-5.8984	163	Vietnamese	1965	L
55	0.832	.9827	O	-4.0553	166	Second Kashmir	1965	L
56	0.131	.6501	I	-.4306	169	Six-day War	1967	H
57	0.408	.8774	I	-.1308	175	Football,(Soccer)	1969	H
58	0.285	.7978	I	-.2259	178	Bangladesh	1971	H
59	0.748	.9741	O	-3.6525	181	Yom Kippur	1973	H
60	1.732	.9998	I	-.0002191	184	Turco-Cypriot	1974	L
61	-0.388	.1285	O	-.1376	190	Ugandan-Tanzanian	1978	L
62	1.245	.9977	I	-.00235	193	Sino-Vietnamese	1979	L

APPENDIX D

DISTRIBUTION

Addressee	No of copies
Deputy Chief of Staff for Operations and Plans Headquarters, Department of the Army ATTN: DAMO-ZA Washington, DC 20310	1
Deputy Chief of Staff for Operations and Plans Headquarters, Department of the Army ATTN: DAMO-ZD Washington, DC 20310	1
Deputy Chief of Staff for Operations and Plans Headquarters, Department of the Army ATTN: DAMO-ZDF Washington, DC 20310	1
Deputy Chief of Staff for Personnel Headquarters, Department of the Army ATTN: DAPE-ZA Washington, DC 20310	1
Deputy Chief of Staff for Personnel Headquarters, Department of the Army ATTN: DAPE-ZBR Washington, DC 20310	1
Deputy Chief of Staff for Logistics Headquarters, Department of the Army ATTN: DALO-ZA Washington, DC 20310	1
Deputy Chief of Staff for Logistics Headquarters, Department of the Army ATTN: DALO-PLF Washington, DC 20310	1

Addressee	No of copies
Commander US Army Logistics Center Fort Lee, VA 23801	1
Deputy Under Secretary of the Army (Operations Research) Washington, DC 20310	2
Chief of Staff, Army ATTN: DACS-DMO Washington, DC 20310	1
Assistant Secretary of the Army (Manpower & Reserve Affairs) Washington, DC 20310	1
Assistant Secretary of the Army (Research, Development, and Acquisition) Washington, DC 20310	1
Comptroller of the Army Headquarters, Department of the Army Washington, DC 20310	1
Director US Army TRADOC Analysis Center White Sands Missile Range, NM 88002	2
Director US Army Materiel Systems Analysis Activity ATTN: AMXSY-LM (Mr. Fox) Aberdeen Proving Ground, MD 21005-5071	2

Addressee	No of copies
Director US Army Ballistic Research Laboratories Building 305 Aberdeen Proving Ground, MD 21005	2
Commander Combined Arms Combat Development Activity Fort Leavenworth, KS 66027	2
Commander Foreign Science and Technology Center 227th Street NE Charlottesville, VA 22901	2
Director Defense Nuclear Agency ATTN: LASS 6801 Telegraph Road Alexandria, VA 20305	2
Commander Army Research Institute 5001 Eisenhower Avenue Alexandria, VA 22333	2
Commander US Army Logistics Evaluation Agency New Cumberland Army Depot New Cumberland, PA 17070	1
Director Defense Logistics Studies Information Exchange US Army Logistics Management Center Fort Lee, VA 23801	1

Addressee	No of copies
Defense Technical Information Center ATTN: DTIC-DDA Cameron Station Alexandria, VA 22314-6145	2
US Army Nuclear and Chemical Agency 7500 Backlick Road, Bldg #2073 ATTN: Library Springfield, VA 22150	2
Commander Air Mobility Research and Development Laboratory Eustis Directorate Fort Eustis, VA 23604	1
Commander US Army Mobility Equipment Research and Development Command Fort Belvoir, VA 22060	1
Commander US Army Research, Development, and Acquisition Information Systems Agency Radford, VA 24141	1
The Pentagon Library (Army Studies Section) ATTN: ANRAL-RS The Pentagon Washington, DC 20310	1
Commander US Army Forces Command ATTN: AFOP-OM Fort McPherson, GA 30330	1

Addressee	No of copies
Director Program Analysis and Evaluation Office of the Secretary of Defense Room 2E330 The Pentagon Washington, DC 20310	2
Organization of the Joint Chiefs of Staff ATTN: J-8 Room 1D936, The Pentagon Washington, DC 20310	2
Joint Deployment Agency Deploying Systems Division MacDill Air Force Base, FL 33608	1
Commandant US Army War College Carlisle Barracks, PA 17013	2
Director Institute for Defense Analyses 1801 North Beauregard Street Alexandria, VA 22311	2
Commander US Army Center of Military History Casimir Pulaski Building 20 Massachusetts Avenue, NW Washington, DC 20314-0200	1

Addressee	No of copies
Commandant US Army War College ATTN: Director, DMSPO Carlisle Barracks, PA 17013	1
Commandant US Army War College ATTN: Library Carlisle Barracks, PA 17013	1
Commandant US Army War College ATTN: CLW Carlisle Barracks, PA 17013	1
Center for Aerospace Doctrine, Research, and Education (AU) ATTN: CADRE/TA Maxwell Air Force Base, AL 36112-5532	1
Air University ATTN: AU CADRE/WGTA (Capt. Taylor) Maxwell Air Force Base, AL 36112-5000	1
Commandant US Air War College Maxwell Air Force Base, AL 36112-5532	1
Commandant US Navy War College Newport , RI 02840	1
President National Defense University ATTN: NDU-LD-CDC Washington, DC 20319-6000	1

Addressee	No of copies
President National Defense University ATTN: NDU-LD-CDC Washington, DC 20319-6000	1
Commandant Armed Forces Staff College Norfolk, VA 23511	1
Commandant US Army Command and General Staff College Fort Leavenworth, KS 66027	2
US Army Command and General Staff College OR Faculty Fort Leavenworth, KS 66027	2
Commandant US Army Command and General Staff College ATTN: Department of Combat Development, Force Development Fort Leavenworth, KS 66027	1
Superintendent United States Military Academy ATTN: Mathematics Department West Point, NY 10996	1
Superintendent United States Military Academy ATTN: Engineering Department West Point, NY 10996	2
Superintendent Naval Postgraduate School Monterey, CA 93940	2

Addressee	No of copies
Naval Postgraduate School ATTN: Department of Operations Research Code 55PY (Professor Parry) Monterey, CA 93940	1
Commandant US Army Infantry School ATTN: ATSH-IVT Fort Benning, GA 31905	1
Commandant US Army Armor School ATTN: ATSB-CD Fort Knox, KY 40121-5215	1
Commandant US Army Field Artillery School Fort Sill, OK 73503	1
Commandant US Army Air Defense School Fort Bliss, TX 79916	1
Commandant US Army Aviation School Fort Rucker, AL 36360	1
Commandant US Army Engineer School Fort Belvoir, VA 22060	1
Commandant US Army Transportation School Fort Eustis, VA 23604	1

Addressee	No of copies
Commandant US Army Intelligence Center and School ATTN: ATSI-TD Fort Huachuca, AZ 85613	1
Commander US Army Ordnance Center and School ATTN: ATSL-CMT Aberdeen Proving Ground, MD 21005-5201	1
Commandant US Army Ordnance, Missile and Munitions Center and School ATTN: ATSK-CMT Redstone Arsenal, AL 35897-6000	1
Commander in Chief United States Readiness Command ATTN: RCDA MacDill Air Force Base, FL 33608	1
Commander US Army Combat Developments Experimentation Command Fort Ord, CA 93941	2
Director US Army Human Engineering Laboratory Aberdeen Proving Ground, MD 21005-5001	1
US Army Human Engineering Laboratory Field Office ATTN: SLCHE-FI 220 7th Street Charlottesville, VA 22901	1

Addressee	No of copies
CINCPAC Staff C3S ASII Box 32A Camp Smith, HI 96861	2
Commander US Army Western Command ATTN: APOP-SPM Fort Shafter, HI 96858-5100	1
SHAPE Technical Centre ATTN: USRADCO (LTC Lewis Davis) APO NY 09159	2
Commander US Army Information Systems Command Fort Belvoir, VA 22060	1
Commander US Army Health Services Command Fort Sam Houston, TX 78234	1
Commander US Army Medical Research and Development Command Fort Detrick, MD 21701	1
Commander Eighth US Army APO San Francisco 96301	2
Commander US Army, Japan ATTN: AJCS APO San Francisco 96343	2

Addressee	No of copies
Deputy Chief of Staff for Intelligence Headquarters, Department of the Army Washington, DC 20310	2
Commander US Army Intelligence and Security Command ATTN: IACS Arlington Hall Station, VA 22212	1
Commander, USAITAC AIAIT-HI, Tech Info Bldg 203, STOP 314 Washington Navy Yard Washington, DC 20374	1
Commander/Director US Army Engineer Studies Center Casey Building, No. 2594 Fort Belvoir, VA 22060	1
Commander US Army Corps of Engineers 20 Massachusetts Avenue, NW Washington, DC 20314-1000	1
Commander in Chief US Army, Europe & Seventh Army ATTN: AEAGF APO New York 09403	2
Commander in Chief US Army, Europe & Seventh Army ATTN: AEAGX-OR (Dr. Leake) APO New York 09403	1

Addressee	No of copies
Commander US Army Training and Doctrine Command ATTN: ATCD-AU Fort Monroe, VA 23651	1
Commander US Army Training and Doctrine Command Fort Monroe, VA 23651	2
Commander US Army Materiel Command 5001 Eisenhower Avenue Alexandria, VA 22333	2
US Army CE Command Program Analysis and Evaluation Systems Analysis Division Fort Monmouth, NJ 07703	2
Air Force Center for Studies and Analyses AFCSA/SAMI Room 1D363, Pentagon Washington, DC 20330-5425	2
Headquarters, US Air Force Office of Worldwide Management of Studies & Analyses ATTN: AF/SAL The Pentagon Washington, DC 20330	2
Headquarters Foreign Technology Division (AFSC) Wright-Patterson Air Force Base, OH 45433	1

Addressee	No of copies
Commander in Chief Military Airlift Command Scott Air Force Base, IL 62225	1
Commander Tactical Air Command TAC Analysis Group Langley Air Force Base, VA 23665	2
Joint Studies Group (TAC) ATTN: OSS Langley Air Force Base, VA 23665	1
Commander USAF Systems Command Andrews Air Force Base Washington, DC 20334	2
Commandant Air Force Institute of Technology ATTN: AFIT-EN Wright-Patterson Air Force Base, OH 45433	1
Headquarters Rome Air Development Center (AFSC) ATTN: COAM (Mr. T. Humiston) Griffiss Air Force Base, NY 13441	1
Center for Wargaming Naval War College ATTN: Naval Warfare Gaming System Code 33 Newport, RI 02840	2

Addressee	No of copies
President Center for Naval Analyses 4401 Ford Avenue Post Office Box 16268 Alexandria, VA 22302-0268	2
Center for Naval Analyses ATTN: Mr. Duncan Love 4401 Ford Avenue Post Office Box 16268 Alexandria, VA 22302-0268	1
Naval Research Laboratory ATTN: Code 5704 4555 Overlook Avenue Washington, DC 20375	1
Chief of Naval Operations ATTN: OP-96 Room 4A526, Pentagon Washington, DC 20350	2
Commander Naval Air Systems Command ATTN: Mission and Effectiveness Analysis Division (AIR-5264) Washington, DC 20361	2
Commander Military Sealift Command Ship Operations Branch 4228 Wisconsin Avenue Washington, DC 20390	1
US Navy Postgraduate School ATTN: CAPT Wayne Highes (USN Ret) Monterey, CA 93943	1

Addressee	No of copies
Commander Naval Sea Systems Command Code 09B5 Washington, DC 20361	1
US Army Liaison Officer Naval Weapons Center China Lake, CA 93555	1
US LNO Supreme Allied Commander Atlantic Norfolk, VA 23511-5100	2
Marine Corps Operations Analysis Group Center for Naval Analyses 4401 Ford Avenue P. O. Box 11280 Alexandria, VA 22302-0268	2
Director USMC Development and Education Center Quantico, VA 22134	2
Brookings Institute ATTN: Dr. Joshua Epstein 1775 Massachusetts Ave, NW Washington, DC 20310	1
Arthur D. Little Technology Resource Center ATTN: Dr. Joseph Engel 600 Maryland Ave, SW Washington, DC	1
Dr. Wilbur Payne 6269 Bandolero El Paso, TX 79912	1

Addressee	No of copies
Mr. Herbert Weiss P. O. Box 2668 Palos Verdes Peninsula, CA 90274	1
Mr. W. Robert Gerber 6060 Haverhill Ct Springfield, VA 22152	1
Prof. Ronald W. Shephard Royal Ordnance Future Systems Group P. O. Box 243 Shrivenham, Swindon, Wilts SN6 8LA England, UK	1
Director Arroyo Center The RAND Corporation 1700 Main Street Santa Monica, CA 90406	2
Mental Health Research Institute University of Michigan ATTN: Prof. J. David Singer Ann Arbor, MI 48109	1
History Department Wayne State University ATTN: Prof. Melvin Small Detroit, MI 48202	1

Internal Distribution:

Unclassified Library	2
Dr. Robert Helmbold	30

GLOSSARY

1. TERMS UNIQUE TO THIS PAPER

a	Parameter (intercept) in a logistic function
b	Parameter (slope) in a logistic function
x	Attacker's surviving personnel strength as of the end of the battle
x0	Attacker's initial personnel strength
y	Defender's surviving personnel strength as of the end of the battle
y0	Defender's initial personnel strength
A	Attacker's fraction of survivors, given by $A = x/x0 = 1 - FX$
ADV	(Defender's) empirical advantage parameter
BD/Pop ratio	Ratio obtained by dividing the opponent's battle deaths per 10,000 prewar population by the initiator's battle deaths per 10,000 prewar population. Defined for wars, rather than for battles.
Cx	Attacker's casualties, i.e., $Cx = x0 - x$
Cy	Defender's casualties, i.e., $Cy = y0 - y$
D	Defender's fraction of survivors, given by $D = y/y0 = 1 - FY$
FER	Fractional exchange ratio, i.e., FX/FY
FX	Attacker's fractional casualties, i.e., $Cx/x0 = 1 - A$
FY	Defender's fractional casualties, i.e., $Cy/y0 = 1 - D$
P(ATKWIN)	Probability that the attacker wins a battle or a war

2. DEFINITIONS

ABS(z)	Absolute value of z
EXP(z)	Exponential function of z, i.e., the constant e raised to the power z
Logistic function	A function of the form

$$f(z) = EXP(a + bz)/(1 + EXP(a + bz)),$$

where a and b are parameters that determine the exact form of the function. Here a is called the intercept and b the slope.

LOG(z)	Natural logarithm of z
SQR(z)	Square root of z
↑	Exponent, i.e., x ↑ p stands for x raised to the power p

DO BATTLES AND WARS HAVE A COMMON RELATIONSHIP BETWEEN CASUALTIES AND VICTORY?	SUMMARY CAA-TP-87-16

THE REASON FOR PERFORMING THIS STUDY was to examine empirically the range of validity of a particular quantitative relation between the probability of victory in battles and the casualties on each side. This relation was discovered in the course of earlier research conducted at the US Army Concepts Analysis Agency (CAA). An empirical finding that the same relationship holds also for operations above the battle or tactical level would substantially strengthen the empirical basis for an inductive generalization that this relation is fundamental in determining victory in combat operations both at and above the tactical level. Since there are no quantitative data bases on combat operations at the campaign level, examining directly whether the relationship between casualties and victory holds for campaigns was not practicable. However, there are some quantitative data bases of wars that can be used for the purpose, although their data are not completely comparable to those for battles. Thus, using war data is a somewhat indirect approach to the study of whether the relation of victory to casualties found by earlier research to hold for battles holds also for campaigns and similar operations at the operational level. However, it was felt that, whatever its shortcomings, this indirect approach was the only currently feasible way to grapple empirically with the issue of whether or not this relationship between casualties and victory applies to combat operations above the tactical level.

THE PRINCIPAL FINDINGS are that this relation between casualties and victory, or some relation similar to it, quite likely does hold for wars as well as for land combat battles. It also appears that the key variables involved have been quite stable from the early 1800s to the present day.

THE MAIN ASSUMPTION is that the available war data are sufficiently error-free to allow at least a rough comparison to be made between them and the land combat battle data.

THE PRINCIPAL LIMITATIONS are two in number. First, not enough war data are available to establish a precisely definitive quantitative estimate of the relevant statistical parameters. Second, the available war data are not fully comparable with that on battles. For example, the war data give the entire national population at the start of the war, while the battle data give those military personnel actually present on the battlefield. Such lack of comparability tends to obscure any similarity of wars and battles regarding the relationship of casualties to victory.

THE SCOPE OF THE WORK is focused on examining whether the relation between casualties and victory in land combat battles, discovered in the course of earlier research, holds also for wars. The paper also includes a brief exploration of the trend over historical time of the key quantities involved, and identifies some issues that would make good topics for future research.

THE STUDY OBJECTIVE was to examine whether the relation between casualties and victory in battle, discovered in the course of earlier research, holds also for wars.

THE STUDY SPONSOR was the US Army Concepts Analysis Agency.

THE STUDY EFFORT was directed by Dr. Robert L. Helmbold, Office of Special Assistant for Model Validation, US Army Concepts Analysis Agency.

COMMENTS AND SUGGESTIONS may be sent to Director, US Army Concepts Analysis Agency, ATTN: CSCA-MV, 8120 Woodmont Avenue, Bethesda, MD 20814-2797.